

***Measuring Teacher Dispositions:
An Application of the Rasch Model to a Complex
Accreditation Requirement***

**Judy R. Wilkerson
William Steve Lang**

University of South Florida St. Petersburg

**IOMW
Cairns, Australia
July 2004**

Measuring Teacher Dispositions:

An Application of the Rasch Model to a Complex Accreditation Requirement

The construct of dispositions is well defined in national standards, and U.S. colleges of education are required to assess candidate dispositions to meet accreditation requirements. Measurement, however, is virtually non-existent. On-line reviews of college accreditation reports indicate that colleges are attempting to assess dispositions without the use of sound measurement techniques or adequate definitions of the construct. The end result, of course, is a reliance on face validity. Rasch measurement provides a much needed solution to scaling the dispositions needed for good teaching. This paper presents early work in the development of three related instruments measuring ten national principles related to dispositions. The instruments use different item structures and response formats, which will eventually be combined into a single disposition scale. Current results are promising for the first two components.

The Problem

U.S. teacher educators are faced with serious challenges to demonstrate the quality of the graduates they prepare. These challenges are expressed in the country's fixation on accountability and testing. We are witnessing the growth of standardized tests and alternative routes to certification as panaceas intended to solve the teacher shortage crisis in most states. National accreditation and state program approval agencies are attempting to stem the tide through the ever-increasing demand for standards-based assessment data documenting teacher quality. The common set of standards used nationally for accreditation was developed by the Council of Chief State School Officers and promulgated by the Interstate New Teacher Assessment and Support Consortium (INTASC) in the form of ten principles. Each of the principles includes indicators written at the knowledge, skill, and dispositional levels, forming constructs that can be measured.

Sadly, neither the profession nor the accreditors have realized the need for objective measurement. This is probably largely a function of what Stiggins (2000) and Popham (2004) bemoan as assessment illiteracy. They are satisfied, at best, with ordinal scales for poorly constructed criteria on ill-defined tasks, or, at worst, with counting papers in portfolios constructed without regard to any form of psychometric consideration (Wilkerson and Lang, 2003). Although not the subject of this paper, the authors are building an ability scale made up of performance tasks to provide a solution to this problem (Wilkerson and Lang, 2004.)

In this paper, however, we are focused on the subset of principle indicators called dispositions. Searches for solutions to assessing dispositions have been fruitless because of the affective nature of the construct and different item types typically used. Teacher educators are clamouring for answers to this requirement since it is being imposed by the National Council for the Accreditation of Teacher Educators (NCATE). NCATE (2002) requires the measurement of dispositions as part of its accreditation requirements for teacher education programs. The first standard, entitled, "Candidate Knowledge, Skills, and Dispositions," requires that:

Candidates preparing to work in schools as teachers or other professional school personnel know and demonstrate the content, pedagogical, and professional knowledge, skills, and dispositions necessary to help all students learn. Assessments indicate that candidates meet professional, state, and institutional standards.

At a workshop these authors conducted last year for the American Association of Colleges of Teacher Education (Wilkerson, et al., 2003), virtually all participants saw dispositions as their burning issue.

The Construct: What Are Dispositions?

We will begin this discussion with two definitions of dispositions:

- Mirriam Webster on-line dictionary: A prevailing tendency, mood, or inclination; temperamental makeup; the tendency of something to act in a certain manner under given circumstances.
- NCATE (2001): The values, commitments, and professional ethics that influence behaviours toward students, families, colleagues, and communities and affect student learning, motivation, and development as well as the educator's own professional growth. Dispositions are guided by beliefs and attitudes related to values, such as caring, fairness, honesty, responsibility, and social justice. For example, they might include a belief that all students can learn, a vision of high and challenging standards, or a commitment to a safe and supportive learning environment.

These definitions help to shed some light on what dispositions are, and we can see intuitively why they are important. Obviously, teacher educators do not want to graduate teachers who do not care about children, are not fair, are dishonest and irresponsible, etc. Many teacher educators, however, end their thinking about these attributes without delving a little more deeply into some specific affective attributes that correlate closely with the knowledge and skills teachers need to have to be effective.

Fortunately, guidance is provided to the community by the common set of national standards developed by the Council of Chief State School Officers and promulgated by the Interstate New Teacher Assessment and Support Consortium (INTASC) in the form of ten principles. Each of the principles includes indicators written at the knowledge, skill, and dispositional levels, forming constructs that colleges are required to measure.

When we begin to conceptualize the INTASC Principles as hierarchical in nature, the need for measuring dispositions is clear. If a teacher learns what elements comprise a good lesson plan and then demonstrates on multiple occasions that he/she has the appropriate level of skill to produce (and hopefully deliver) effective lesson plans, we are often lulled into believing that our job is done. They have the knowledge and can apply it, but what happens if they do not think it is important. No pre-graduation faculty assessment of "proficient in planning" will ever compensate for the damage that can be done by the teacher who thinks lesson planning is a boring waste of time. That teacher will just stand up and deliver. And that is why dispositions are, in the long run, more important than knowledge and skills.

The INTASC Principles lay the foundation upon which we can build solid assessment devices for measuring teacher dispositions. Take for example the three indicators provided in INTASC Principle #7 on planning:

- The teacher knows when and how to adjust plans based on student responses and other contingencies.(Knowledge)

Measuring Teacher Dispositions

- The teacher believes that plans must always be open to adjustment and revision based on student needs and changing circumstances.(Dispositions)
- The teacher responds to unanticipated sources of input, evaluates plans in relation to short- and long-range goals, and systematically adjusts plans to meet student needs and enhance learning.(Skills)

The teacher knows about it, believes in it, and does it. Traditionally, we assess knowledge using tests and measures with considerable practice and confidence in the measures. It is difficult to determine if the teacher believes in something enough to do it on his/her own and plan for it when no one is watching. But if we do not attempt to project whether the skills will continue to be applied in the “real” world, we have partially failed in our obligation to produce highly qualified teachers, leaving no child behind. Therein lays the challenge.

The Importance of Inference in Measuring Dispositions

In general, levels of inference are dictated by how hard it is to score an instrument. If a machine or a relatively untrained rater can score with a high degree of accuracy, then the level of inference is low. As the level of inference increases, the difficulty of scoring and the need for rater training, rubrics, and examples increase,. It becomes more difficult to make the judgment about the observed response. In the development of this scale, we are using three instruments of increasing levels of inference: a Thurstone scale, a teacher questionnaire, and an interview (focus group) of a sample of K-12 students.

Method

As mentioned above, there are three instruments used in development of this scale, each requiring a higher level of inference. At the lowest level is a Thurstone scale, which requires the respondent to simply agree or disagree with 50 statements. This scale can be machine-scored. Items were constructed with the intention of varying levels of difficulty among each of the INTASC Principles. An example of two items, aligned with one of the INTASC indicators, follows:

INTASC Principle	Thurstone Statements
3.4: The teacher is sensitive to community and cultural norms.	Agree: I believe good teachers learn about the students' backgrounds and community so they can understand students' motivations. Disagree: I prefer to live in one community and teach in a different one because I don't really understand the values of many of the students.

Questionnaires and interviews are more difficult to score and a more difficult to guess on, so they provide the next level of useful assessment of dispositions. Unlike an agree/disagree scale, the respondents do not have a 50% chance of getting it right. The introduction of a judge, though, adds the possibility of a rater effect.

For questionnaires and interviews, we can develop rubrics and anticipate likely responses. Raters need some training and some examples, but because of the potential to anticipate responses, particularly on short answer questionnaires, the difficulty associated with analysis is not extreme. In this series, we are using a teacher questionnaire composed of nine items, each of which is comprised of a sub-set of questions targeting a specific INTASC indicator. Judgments are made, based upon what the teacher writes, about the INTASC Principle being assessed, on a three point scale of “target,” “acceptable”, or “unacceptable.” An example for this method, using another INTASC principle, follows:

INTASC Principle	Questionnaire Item
1.1: The teacher realizes that subject matter knowledge is not a fixed body of facts but is complex and ever-evolving. S/he seeks to keep abreast of new ideas and understandings in the field.	How have you kept abreast of current developments in your field? For example, did you attend any workshops, subscribe to any journals, read or buy a new book? If so, describe in one to two sentences something you learned and the source.

When we look at the results, there are answers that clearly show differences in values. Take, for example, the following two responses, where the first one provides evidence of a teacher who believes in learning strongly enough to actively engage in reading outside the college requirements while the other just counts on faculty to tell her everything she needs to know.

- I am a member of the National Council for Exceptional Children and receive journals from them. My roommate is a member of the NEA and receives a journal from them that I read. I also daily check the CNN website under EDUCATION news for any and all happenings around the country, especially since when I graduate I will be teaching in another state. I not only have started a classroom library for my future students but I have bought any books that have been recommended to me by my peers or supervisors and professors. Examples would be *The First Days of School, Lesson Plans for Eric Carle Books, Love and Logic, the Essential 55, and Educating ESME*. I have learned so many things and cannot describe any in one or two sentences. I think most importantly I have learned that no only will I be a teacher but I will continually be a student. I must be in order to give my students the best education that they deserve.
- I am only aware of developments in my field through school. What I have learned in school keeps me updated on what is going on in the school system.

There is still some significant chance, however, that the answers can be faked when we ask only the teacher or teacher candidate to respond. So, at the next level of inference are focus groups of K-12 students and observations of the teacher. There is no substitute for first hand observation of a teacher's performance or for hearing what children have to say about their teacher. These methods, though, are more complex to analyse than simple questionnaires, often because there is interaction among the group members and conflicting evidence. While faking becomes difficult at this point, there are trade-offs with the involvement of other children. Judgement has to be applied to sort good data from noise. The challenge to the rater here is to determine whether or not extraneous factors, like a bad day or a recalcitrant child, make the results muddy.

In this series of instruments we have organized eight sets of questions around the activities being conducted in a classroom so that small focus groups of five students can respond within a familiar context. Their answers are recorded by the interviewer on a form, with judgments made about the INTASC Principle being assessed, again on a three point scale of "target" "acceptable", or "unacceptable." An example follows:

INTASC Principle	Focus Group Questions
5.2: The teacher understands how participation supports commitment, and is committed to the expression and use of democratic values in the	<p><u>Group work:</u> (1.2, 5.2)</p> <ul style="list-style-type: none"> • Usually, when you work in groups, do group members tend to work alone and compile the work at the end, or do they tend to complete most/all components together? Does the teacher do anything to

Measuring Teacher Dispositions

classroom.	<ul style="list-style-type: none">ensure that students work together? If so, what does he/she do?When your groups do their work, do they attempt to reach consensus on group operations and products, or does one person tend to dominate? What does your teacher do if someone dominates the group?
------------	---

When there are clear patterns of concern among children in a group and comments tend to reflect similar concerns, we can infer a problem may exist with the teacher's dispositions. Comments such as the following provide an example for one teacher candidate:

- I think that the smart people get most of the attention. The dumber students don't get talked to as much as the smart ones.
- We usually work altogether but some kids think they are smarter and just work by themselves.

Validity and Reliability

Evidence of validity and reliability are particularly important in measuring dispositions because of the newly recognized need to measure this construct. If one uses the INTASC Principles as the basis for designing instruments, evidence of construct validity starts from item alignment with the defined concepts. Creating a blueprint that ensures coverage of most or all of the dispositional statements in one form or another helps to provide evidence of content validity. While techniques are available to obtain evidence of reliability, the different item types make internal consistency more complex. The Rasch model of Item Response Theory is promising for quantitative analysis of this disposition scale; since it not only estimates reliability but also provides additional evidence of construct validity when measures perform as expected.

The Rasch model was chosen because it is a proven measure construction method with different item structures such as dichotomous response, rating scales, partial credit, Poisson counts, and Bernoulli trials (Stone, 2004). It also creates easily presentable rulers of interval level data, helps to interpret judge bias and rater effects, and diagnoses person and item fit to the construct measured (Bond & Fox, 2003).

Results

At present, we have analysed data on 486 respondents on the belief scale. A sample of 48 examinees completed both the belief scale and the questionnaire. Analysis was conducted using Winsteps software (Linacre, 2003).

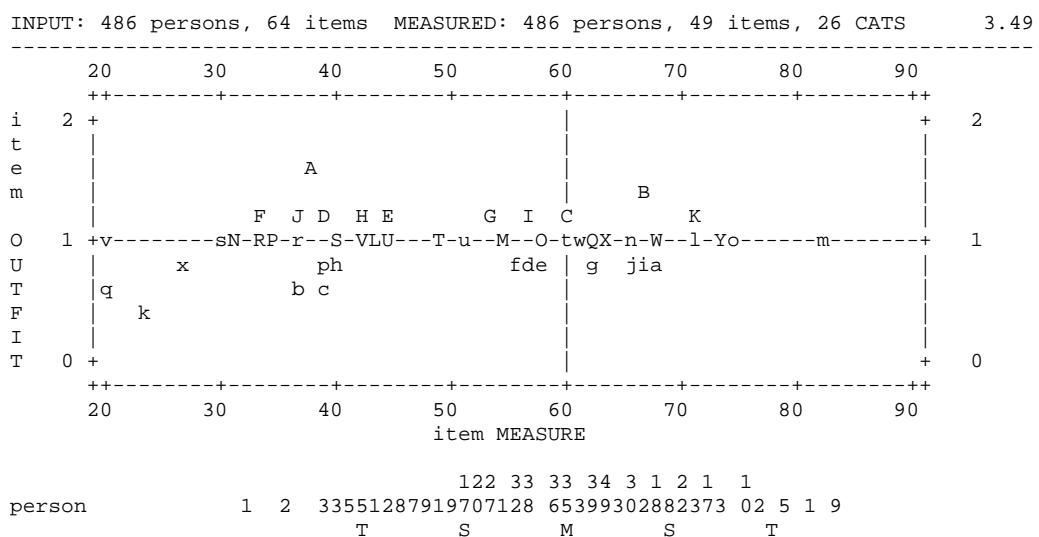
Below we have reproduced several tables from Winsteps output illustrating the combined calibration of belief scale items and questionnaire items combined to measure the construct of INTASC disposition principles. The final run removed misfitting items.

Table 1 Misfitting Items Removed After First Analysis

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT		OUTFIT		PTMEA		items	G
					MNSQ	ZSTD	MNSQ	ZSTD	CORR.	DISPLACE		
4 ;	440	486	35.28	1.59	1.12	1.0	1.92	4.5	-.09	.00	INTASC 02	2
32 ;	355	486	48.82	1.07	1.19	3.5	1.44	5.8	-.02	.01	INTASC 02.1	2
5 ;	451	486	32.16	1.79	1.05	.4	1.58	2.6	.02	.00	INTASC 10.5	2
30 ;	359	485	48.25	1.08	1.14	2.6	1.42	5.3	.03	.01	INTASC 05	2
17 ;	433	486	36.94	1.50	1.11	1.0	1.35	2.2	.04	.00	INTASC 09	2
7 ;	460	486	28.87	2.05	1.05	.4	1.29	1.2	.04	.00	INTASC 08.1	2
31 ;	356	483	48.42	1.08	1.15	2.8	1.20	2.7	.08	.01	INTASC 05	2
37 ;	292	484	55.21	.98	1.16	4.8	1.16	3.8	.09	.01	INTASC 07	2
2 ;	458	486	27.49	2.00	1.02	.2	1.58	2.0	A .10		INTASC 09.2	2

A combination of large misfit, low point-biserial correlations, and extreme difficulty (too easy in the case of item number 2) were examined along with the principal components analysis. Nine of the original 58 items were felt to require removal or revision. The resulting scale illustrates in Table 2 an overall variability within the values expected for the probabilistic model (Smith, 2003).

Table 2 Plot of Scale with Poorer Items Removed



Measuring Teacher Dispositions

Table 3 below illustrates the item statistics of the final scale.

Table 3 Items Included After Poorer Items Removed.

INPUT: 486 persons, 64 items MEASURED: 486 persons, 49 items, 26 CATS 3.49

person: REAL SEP.: 1.72 REL.: .75 ... item: REAL SEP.: 8.95 REL.: .99
items STATISTICS: MISFIT ORDER

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	ERROR	INFIT		OUTFIT		PTMEA	DISPLACE	items	G
					MNSQ	ZSTD	MNSQ	ZSTD				
12	424	486	38.13	1.42	.90	-1.0	1.55	3.1	A .27	.01	INTASC 08	D
48	173	484	66.76	1.02	1.19	4.6	1.32	5.0	B .14	.01	INTASC 04	D
43	245	486	59.79	.98	1.19	5.5	1.27	5.6	C .16	.01	INTASC 06	D
18	415	482	39.30	1.38	.88	-1.3	1.25	1.7	D .33	.01	INTASC 10.5	D
28	379	482	44.98	1.18	1.16	2.4	1.22	2.1	E .15	.01	INTASC 10	D
13	443	484	33.16	1.69	1.03	.3	1.19	.9	F .18	.00	INTASC 08.1	D
35	309	480	53.10	1.02	1.18	4.3	1.16	2.5	G .18	.01	INTASC 03.1	D
25	400	483	42.05	1.27	1.12	1.6	1.17	1.3	H .16	.01	INTASC 02	D
40	280	483	56.24	.99	1.16	4.4	1.15	3.0	I .21	.01	INTASC 10.2	D
15	429	486	37.09	1.47	1.04	.5	1.11	.7	J .21	.00	INTASC 05.4	D
49	135	484	70.97	1.08	.97	-.6	1.11	1.4	K .35	.01	INTASC 01	D
26	395	484	42.90	1.24	1.11	1.4	1.07	.6	L .21	.01	INTASC 10	D
34	298	481	54.35	1.01	1.08	2.2	1.09	1.7	M .27	.01	INTASC 05.4	D
10	448	484	31.65	1.79	1.04	.3	1.09	.5	N .18	.00	INTASC 01	D
38	260	484	58.20	.98	1.06	1.9	1.09	1.8	O .30	.01	INTASC 03.1	D
14	441	486	34.22	1.62	1.09	.7	1.03	.2	P .17	.00	INTASC 08	D
44	220	485	62.12	.98	1.05	1.5	1.08	1.8	Q .31	.01	INTASC 08.2	D
11	444	486	33.41	1.67	1.04	.4	1.07	.4	R .19	.00	INTASC 07	D
24	412	486	40.40	1.33	1.02	.3	1.04	.3	S .26	.01	INTASC 09.1	D
29	346	484	49.36	1.08	1.01	.2	1.03	.4	T .33	.01	INTASC 04.1	D
22	395	485	43.10	1.23	.92	-1.1	1.01	.2	U .36	.01	INTASC 03	D
27	390	485	43.81	1.21	.93	-1.0	1.00	.1	V .36	.01	INTASC 03.2	D
55	40	39	67.23	2.56	1.00	.1	1.00	.1	W .32	.05	INTASC 7.1, 7.3	0
58	48	39	63.49	2.24	1.00	.0	1.00	.1	X .35	.05	INTASC 10.3, 10.5	0
50	112	478	73.61	1.15	1.00	.0	.97	-.3	Y .34	.01	INTASC 05.4	D
8	459	483	27.09	2.14	.99	.0	.85	-.5	z .22	.00	INTASC 03.1	D
45	214	485	62.76	.98	.94	-1.9	.98	-.4	w .43	.01	INTASC 02.1	D
1	470	483	20.52	2.85	.94	-.1	.98	.1	v .19	.00	INTASC 10.2	D
33	327	486	51.64	1.04	.98	-.5	.96	-.5	u .38	.01	INTASC 04	D
42	232	477	60.57	.99	.95	-1.5	.98	-.5	t .41	.01	INTASC 01.2	D
9	456	486	29.54	1.94	.97	-.2	.90	-.3	s .24	.00	INTASC 01.2	D
16	427	484	37.11	1.47	.86	-1.4	.96	-.2	r .38	.01	INTASC 01	D
3	470	483	20.48	2.85	.94	-.2	.70	-.7	q .24	.00	INTASC 03.4	D
21	421	484	38.37	1.41	.93	-.7	.90	-.6	p .34	.01	INTASC 06.2	D
57	33	39	72.91	2.89	.93	-.2	.93	-.2	o .39	.05	INTASC 9.4	0
47	187	484	65.32	1.00	.93	-2.0	.93	-1.3	n .44	.01	INTASC 03.4	D
51	26	39	82.04	3.17	.92	-.4	.93	-.3	m .40	.05	INTASC 01.1	0
52	34	39	70.91	2.48	.92	-.4	.92	-.4	l .44	.05	INTASC 01.4	0
6	469	486	23.30	2.51	.91	-.3	.46	-1.9	k .30	.00	INTASC 03	D
53	43	39	65.36	2.52	.89	-.5	.89	-.5	j .47	.05	INTASC 05.1	0
54	41	39	66.73	2.39	.89	-.6	.89	-.6	i .47	.05	INTASC 05.3	0
20	414	486	40.04	1.34	.87	-1.5	.89	-.8	h .40	.01	INTASC 09	D
46	214	481	62.56	.99	.89	-3.5	.88	-2.7	g .49	.01	INTASC 02	D
39	285	485	55.91	.99	.88	-3.5	.84	-3.4	f .50	.01	INTASC 03.4	D
41	261	485	58.18	.98	.88	-3.9	.84	-3.7	e .50	.01	INTASC 04.1	D
36	287	486	55.74	.99	.87	-3.8	.85	-3.1	d .50	.01	INTASC 06	D
23	412	479	39.26	1.38	.86	-1.6	.62	-3.0	c .46	.01	INTASC 04	D
19	429	486	37.09	1.47	.84	-1.7	.63	-2.6	b .45	.00	INTASC 02.1	D
56	40	39	67.17	2.71	.73	-1.4	.73	-1.4	a .71	.05	INTASC 8.1	0

Table 4 provides summary statistics for the resulting scale. Person reliability is satisfactory at .75 with a separation of 1.72. Item reliability and separation are .99 and 8.95 respectively. The expected values for the outfit means and standard deviation if the data fit the model are 0.0 and 1.0 respectively. This final version of the scale shows reasonable person values of $M = .0$ and $SD = 1.1$. The item values of $M = .1$ and $SD = 1.9$ illustrate some outlier sensitivity, but this may be a result of the scale combination of two item types with the questionnaire type clearly an extreme compared with the Thurstone type. The Real Reliability is a lower bound statistic while the model is an upper bound. In this case, the range is minimal.

Table 4 Summary Statistics of Final Scale

INPUT: 486 persons, 64 items MEASURED: 486 persons, 49 items, 26 CATS 3.49

SUMMARY OF 486 MEASURED persons							
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	OUTFIT MNSQ	ZSTD
MEAN	29.9	41.5	60.01	4.21	.99	-.1	.99
S.D.	6.1	2.5	8.82	.67	.24	1.2	.55
MAX.	50.0	49.0	83.62	7.58	2.01	4.6	3.64
MIN.	11.0	29.0	32.49	3.31	.54	-2.9	.24
REAL RMSE	4.43	ADJ.SD	7.63	SEPARATION	1.72	person RELIABILITY	.75
MODEL RMSE	4.26	ADJ.SD	7.72	SEPARATION	1.81	person RELIABILITY	.77
S.E. OF person MEAN	= .40						
VALID RESPONSES: 84.6%							
SUMMARY OF 49 MEASURED items							
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	OUTFIT MNSQ	ZSTD
MEAN	296.6	411.2	50.00	1.57	.98	-.1	.99
S.D.	147.6	164.4	15.45	.65	.10	2.0	.19
MAX.	470.0	486.0	82.04	3.17	1.19	5.5	1.55
MIN.	26.0	39.0	20.48	.98	.73	-3.9	.46
REAL RMSE	1.72	ADJ.SD	15.36	SEPARATION	8.95	item RELIABILITY	.99
MODEL RMSE	1.70	ADJ.SD	15.36	SEPARATION	9.02	item RELIABILITY	.99
S.E. OF item MEAN	= 2.23						

Table 6 is the logistic ruler produced by the data. It shows a normal distribution of persons, which is expected in dispositions -- unlike skills. With skills, we expect the majority of persons in teacher training to have mastered the skills assessed, whereas, dispositions, which are typically not taught, tend to spread into a more normal pattern. Note also that there is a wide range in the measures for items, and they are equally spread throughout the scale, showing minimal gaps in our coverage of the construct. This is even clearer in Table 7, which identifies the INTASC Principles measured by number.

Here, an X refers to a higher inference Questionnaire item while a D refers to a lesser inference dichotomous item. It's expected that knowledge of the "right" disposition may not transfer into appropriate behaviour. The next item type to be added will be a focus group of perception of students of the disposition of the teacher. As the construct is envisioned, the scale item types should "stack" as each inference step is necessary for the next. This initial scale confirms the construct item locations.

Measuring Teacher Dispositions

Table 6 Logistic Ruler of the Scale Indicating Item Type

INPUT: 486 persons, 64 items		MEASURED: 486 persons, 49 items, 26 CATS	3.49
<hr/>			
	MAP OF persons AND items		
MEASURE			MEASURE
<more>	----- persons	----- items	<rare>
110	+		110
100	+		100
90	+		90
80	. #####	T X	80
	.	+	
	. ##	T	
	#		
	#####		
	.	# XD	
	#####	XD	
70	#####	+	70
	#####	S	
	#	XXXD	
	#####	S XD	
	#####	XDD	
	#####	D	
60	#####	M+ DD	60
	#####	DD	
	#####	DD	
	#####	DD	
	#####	D	
50	#####	S D	50
	.	#####+M D	
	####		
	.		
	###	D	
	###	DDD	
40	###	D	40
	.	+ DDDD	
	.	DD	
	#	DDD	
	#	S D	
	.	DD	
30	.	D	30
		D	
		D	
20		+ DD	20
<less> ----- persons ----- items -----<frequent>			
EACH '#' IN THE person COLUMN IS 2 persons; EACH '.' IS 1			

X=Questionnaire Item

D=Belief Scale Item

Table 6 Logistic Ruler of the Scale Indicating INTASC Principle

INPUT: 486 persons, 64 items MEASURED: 486 persons, 49 items, 26 CATS 3.49

80	persons MAP OF items	
	<more> <rare>	
	.##	
	T INTASC 01.1	
	. +	
	. # T	
	.	
	.##	
	INTASC 05.4 INTASC 9.4	
	INTASC 01 INTASC 01.4	
70	#####+	
	.#####+ S	
	INTASC 04 INTASC 05.3 INTASC 7.1,7 INTASC 8.1	
	S INTASC 03.4 INTASC 05.1	
	INTASC 02 INTASC 02.1 INTASC 10.3,	
	INTASC 08.2	
60	######+	
	.######+ M+ INTASC 01.2 INTASC 06	
	.######+ INTASC 03.1 INTASC 04.1	
	.######+ INTASC 03.4 INTASC 10.2	
	.######+ INTASC 05.4 INTASC 06	
	.######+ INTASC 03.1	
	.######+ S INTASC 04	
50	.## +M INTASC 04.1	
	.##	
	. #	
	## INTASC 10	
	## INTASC 03 INTASC 03.2 INTASC 10	
	. # T INTASC 02	
40	. + INTASC 04 INTASC 09 INTASC 09.1 INTASC 10.5	
	. INTASC 06.2 INTASC 08	
	. INTASC 01 INTASC 02.1 INTASC 05.4	
	. S INTASC 08	
	. INTASC 07 INTASC 08.1	
	. INTASC 01	
30	. + INTASC 01.2	
	INTASC 03.1	
	INTASC 03	
20	. + INTASC 03.4 INTASC 10.2	
	<less> <freq>	
EACH '#' IS 4.		

In our initial use of these instruments, we have found some remarkable results. For example, there are appropriate ranges of logit scores (the ruler was arbitrarily rescaled from 0 to 100) within INTASC Principle items, and the order is logically sound. Items we expected to be more difficult were more difficult, as can be seen in the following set of items that range from 30 to 73 for items on a single INTASC Principle (critical thinking), indicating that the teachers measured can provide a socially correct response about students needing to learn to think but those less committed to critical and creative thinking believe it should occur outside their own classrooms in the fine arts.

- **Agree:** 23. Students need to learn to think, and that is a goal that I have that is built into all my lessons. (least difficult –scale value=30)
- **Disagree:** 41. Students who can't think are basically dumb, so I don't think giving kids time to "brainstorm" is anything except a waste of time. (second most difficult = scale value=41)
- **Agree:** 29. It's more important that the students learn to think and be creative than it is that they know the material covered by the lessons. (third most difficult –scale value=59)
- **Disagree:** 33. The subject I teach doesn't focus on creativity or thinking skills, but I believe all students should be exposed to art and music while in school. (most difficult - scale value= 73)

In addition to the results presented above supporting the unidimensionality of the disposition construct, correlation of person ability scores with GPA yields a result of $r=.20$ for this sample. Intuitively we know that some high achievers have attitude deficits, and some low achievers are warm and fuzzy beings. This result supports the lack of relationship between knowledge/skill and disposition, thereby providing strong support for the need to measure dispositions as a separate and equally important construct.

We also think that the order of item measure types is somewhat as expected. As the inference of the scale moves from the dichotomous belief scale to the partial credit questionnaire, the item structure is building more "difficulty" into the disposition principle. If we are correct, the addition of even higher inference assessments (Focus Group, Observational, Projective) will continue to build the calibrated pool of items for the disposition construct using additional item structures.

Conclusions

We realize that we have a long way to go before this work is done. The belief scale is clearly more developed than the other instruments. We recently attended a workshop in our University from a consultant being promoted by NCATE as a measurer of dispositions. He advocated the use of a rubric with three proficiency levels. By way of example, we offer the high point on the scale, called "optimal."

Candidates show clear and consistent evidence of an orientation toward continuous, self-motivated inquiry aimed at professional learning and development. In addition to reading a variety of professionally related periodical literature, candidates read professional books not required for school or work, or participate in collaborative literature circles or study groups focusing on professional topics. Candidates attend state, regional or national professional conferences or other training opportunities. Candidates conduct

classroom-based action research to inform their practice. They invite observation of their own teaching by others.

Somewhat apologetically he announced his background in early childhood education and lack of psychometric training. He proceeded to present his assessment solution to get the data needed for his rubric – a “Disposition Evaluation Form” with dichotomous ratings of observed or not observed, administered at entry to student teaching. Item number one was:

Regularly establishes professional development goals, takes action to attain those goals, and assesses the outcomes of action they have taken.”

Although no data collection methodology was presented, the audience “oohed” and “aahed” appropriately. No one asked how this was to be observed. Of course, the presenter mentioned words like validity and reliability, and a hush fell over the room. Without measurement models that will make sense of intuitively different item types, measurement of INTASC’s disposition principles will resort to assessment with no validity and reliability.

References

Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*, Mahwah, NJ, Lawrence Erlbaum.

Council of Chief State School Officers (1992). INTASC's *Model Standards for Beginning Teacher Licensing, Assessment and Development: A Resource for State Dialogue*. Washington, D.C.: Author. Retrieved Sept. 15, 2004 from

http://www.ccs.org/projects/Interstate_New_Teacher_Assessment_and_Support_Consortium/

National Council for Accreditation of Teacher Education (2001). *Professional Standards for the Accreditation of Schools, Colleges, and Departments of Education*. Washington, D.C.: Author.

Popham, J. (2004). *All About Accountability: Why Assessment Illiteracy is Professional Suicide*. Educational Leadership (62)1.

http://www.ascd.org/cms/objectlib/ascdframeset/index.cfm?publication=http://www.ascd.org/publications/ed_lead/200409/popham.html

Stiggins, R. (2000). *Specifications for a Performance-Based Assessment System for Teacher Preparation*. National Council for Accreditation of Teacher Education, Washington, D.C. Retrieved June 15, 2004 from

<http://www.ncate.org/resources/commissioned%20papers/stiggins.pdf>

Stone, G. E. (2001). Understanding Rasch Measurement: Objective Standard Setting (or Truth in Advertising). *Journal of Applied Measurement*, 2, 2, 2002.

Wilkerson, J.R., & Lang, W.S. (2003). Portfolios, the Pied Piper of teacher certification assessments: Legal and psychometric issues. *Education Policy Analysis Archives*, 11:45. Retrieved December 20, 2003 from <http://epaa.asu.edu/epaa/v11n45/>.

Wilkerson, J., Lang, W.S. (2004)). Designing Standards-Based Tasks and Scoring Instruments to Collect and Analyze Data for Decision-Making. Workshop for annual meeting of the American Association of Teacher Educators in Washington, D.C.